The Formation of Natural Concepts

R. T. Kellogg and L. E. Bourne, Jr.

University of Colorado

Institute for the Study of Intellectual Behavior

Report No. 62

April, 1977

Abstract

The literature pertaining to models of pattern recognition was reviewed with an emphasis on assessing the adequacy of the major approachs in accounting for real world scene analysis. It was argued that (a) a relational model is required to account for the problems engendered by context in complex scenes, (b) a feature frequency principle accounts for differences in typicality among exemplars of a category, (c) data taken to support a family resemblance principle of categorization are better interpreted in terms of a structural-frequency theory.

The Formation of Natural Concepts

Not all instances of real world concepts are equally good examples of the core meaning of the concept. Thus, reaction time to verify that _X_ is a _Y_ is an inverse function of the degree to which X is considered typical of the concept Y (e.g., Rosch, 1973). As Rosch (1973) points out, traditional definitions of concepts in terms of _absolute_ criterional features and relations (Bourne, 1970) are unable to account for gradients of membership found within ecological concepts.

One goal of the present paper is to review theories that do encompass both a person't ability to classify correctly and to organize real world concepts according to typicality. The relevant alternatives can be described with respect to two central questions about the content of mental representations: Does hte mental representation of a concept include independent or structurally related features? and Is degree of criteriality of a feature represented in terms of distance or frequency information?

A second goal of this paper is to argue that a structural-frequency point of view is the most promising alternative as a description of human categorization. A review of the literature will reveal the appropriateness of feature frequency as an explanation of category membership gradients. In our theory, every feature which appears among instances of a concept is considered part of the definition of concept. Features are not labeled as relevant or irrelevant, however. Rather, feature criteriality is viewed as a continuum. One or more features may be sufficient for describing the boundary of a category organization. We assume that criteriality is a direct function of the frequency distribution of features on each constituent attribute or dimension.

Another task undertaken here is to extend the structural-frequency theory to some recently reported data that have been taken as evidence for a family resemblance principle in human categorization (Rosch & Mervis, 1975; Rosch, Simpson, & Miller, 1976). The family resemblance principle, as we shall see, is closely related to feature frequency; consequently, empirical demonstrations said to support one theory can often be interpreted in light of the other theory. Both theories operate in the context of a feature format and, although quite different in theoretical ways, the predictive model used by Rosch in the laboratory is identical to a feature frequency model (Neumann, Note 1).

In AI work, tractable feature space provides the grist for the classification algorithm. The chief obstacle facing any feature theory of concept representation is the problem of specifying the appropriate features. The problem is particularly severe in the analysis of complex, three-dimensional scenes, such as real world events which an adequate theory of human categorization must address eventually.

Prototype and Feature Representations

Rosch (1973; 1975a) is the strongest proponent of the prototype format point of view. She argues that both the structure (format) and content of mental representations can be conceived of in terms of prototypes or clearest cases. Other instances of a concept "surround" the best example according to the degree to which they are similar to the prototype. Rosch (1973) presented evidence that both perceptual (color categories) and semantic concepts are best described in terms of a prototype representation. The organization of color categories appears to be physiologically determined and universal across cultures. Rosch (1973a) found that a tribe of stone age people, the Dani, demonstrated a preference for members of the color space that are considered prototypical by people of Western cultures. Rosch's result is interesting since the subjects in her learning experiments did not linguistically divide the color space as English speakers do. The Dani possess only two color terms which linguistically divided the space on the basis of brightness rather than hue. That semantic representations also are based on prototypes was supported by a semantic verification experiment (Rosch, 1973). In that study, reaction time to verify that *X* is a member of concept *Y* was inversly related to the degree to which *X* is considered a good member of concept *Y*.

We agree that the *content* of mental representations must reflect the fact that perceptual and semantic categories are internally organized. It is important to draw a distinction, however, between perceptual and semantic categories with respect to the *format* of the representation (see Kieras, 1976). Rosch's more recent work seems to reflect the same sentiment. For instance, Rosch, Simpson, and Miller (1976) propose a distance model of categorization based on a quantitative, Euclidean feature space and a family resemblance principle based on qualitative features of a concept. That is, although gradients of membership were claimed to be part of the *content* of semantic category representations, a feature *format* was assumed.

Color categories may not be amenable to the feature approach; certainly there is evidence that semantic and color categories are different in character. Consider the fact that responses in a same-different task (category same instructions) were faster when the category name was used as a prime than when no prime was goven (Priming effect) to the extent that the items judged same were good examples of the category (Goodness of example effect). For color categories, the Priming X Goodness of Example interaction persisted despite massive rehearsal distributed over a period of three weeks (Rosch, 1973), while the effect disappeared for semantic categories (Rosch, 1975). It is probably best to treat the question of how color concepts are represented separately from the analogous issue with respect to semantic concepts.

Rosch's work was summarized to represent the prototype point of view since it is clear in her writing that both the format and content of the representation is characterized in terms of prototypes, at least in the case of color categories. In contrast, it is not clear that the format of a prototype-transformation viewpoint must be non-featural in nature. Franks and Bransford (1971) claim that the best example of the concept (analog or proposition) is represented in memory along with the transformation rules which generate other instances of the concept by operating on the prototype. To the extent that the transformation rules are a grammar which specifies how features and relations can be combined to form new features, relations and objects, an analog format is less feasible than a feature format.

We would like to suggest that both the analog and the featural viewpoints make the claim that knowledge about the typicality of exemplars is represented. However, with respect to the *format* issue, the analog-prototype view is quite different from feature theories. The former considers the format of the representation to be like a concrete image (cf. Rosch, 1975) or reinstatement of sensory events while the latter is based on an abstract, propositional format (c.f. Winston, 1970).

Independent or Structurally Related Features

The computer metaphor provides a rich vocabulary for discussing human pattern recognition and concept formation. The surge of interest in machine learning during the late fifties and early sixties, as well as more recent attempts at computer scene analysis offer several explicit models of human categorization. Both Hunt (1975) and Duda and Hart (1973) provide excellent reviews of this literature.

A statement of the artificial intelligence approach may clarify the connection between machine models and a feature theory of human concept representation. The *physical* world is conceived as being of infinite dimensionality, which through some process of transduction is reduced to a *pattern space* of finite dimensionality, R. With the "human machine", the pattern space might be thought of as retinal stimulation, a low level, on-off code of cell firings. The pattern space is then generally reduced to some tractable number of features, N. The resulting reduction of dimensionality, N/R, is an important determinant of the complexity of the next step -- the decision algorithm which divides the feature space into K classes. If the set of features neatly divides the categories, then the resulting decision rule is trivial. If the feature space has a large number of dimensions or if the features yield a messy, difficult to separate space, then the decision rule must be more complex if it is to be useful. The psychological problem of visual concept formation is faced with both a large number of dimensions and a feature space in which certain features appear in more than one class.

The above conceptualization, presented in more detail by Andrews (1972), offers a clear view of one major differentiation among types of models--statistical classification models concentrate on the categorization rule, while descriptive models focus on a grammar of the feature space. Consider each in turn.

An assumption of statistical models is that each stimulus object or point in the pattern space can be represented either by a set of measurements defining the axes of some Euclidean feature space or by an ordered list of discrete features on qualitative dimensions of variance. In the case of quantitative, Euclidean features, one can speak of the distance from one pattern to another, while qualitative feature values are not amenable to a distance interpretation since the dimension values do not necessarily form an interval scale (Hunt, 1975). In both cases, however, classification proceeds in essentially the same manner. The principle can be understood by considering a parallel machine (Hunt, 1975) or a simple perceptron (Minsky & Papert, 1969).

Perceptrons.

To start the classification process a perceptron determines or is supplied with the value of each stimulus dimension of the test pattern. This set of $\underline{n}$ measurements may represent a point in Euclidean space or a feature vector. Each feature detector is called a partial predicate. Weights are assigned to each measurement and a weighted sum is taken. A learning algorithm adjusts these weights so that the value of the weighted sum correctly assigns all test patterns to the correct category. Notice that this procedure defines a linear boundary between the categories by combining the evidence from each partial predicate computed in parallel. The weight assigned to each partial predicate is a coefficient of a linear discriminant function (Nilsson, 1965). That the pattern space can be defined in terms of a set of _n_ measures is assumed by all models. If in addition one can specify the nature of the distribution of values on each dimension of measurement, then a parametric model, such as Bayes Theorem, may be used to separate the pattern space appropriately. If assumptions regarding the distributions cannot be made, then non-parametric techniques are appropriate. Included in this category are the nearest neighbor or proximity algorithm, prototype-distance, and cue validity models reviewed by Reed (1973).

Reed (1972) compared these non-parametric models in a series of three experiments to see which best predicted how subjects classified Brunswik faces. The subject first studied two categories of five faces each. The faces varied on four dimensions--length of nose, distance between eyes, height of forehead, and distance between mouth and tip of chin. The data strongly suggested that subjects calculated, in some unspecified manner, a prototype exemplar, which exhibited the mean values on each dimension of variance. Subjects presumably used the prototype of each category to classify test patterns by further calculating the distance between the pattern and the two prototypes. The classification algorithm assigned the pattern to the "closest" category. Note that prototype-distance models define linear boundaries between categories by summing together independently calculated weighted measurements. All minimum distance classifiers are parallel machines (Nilsson, 1965).

A cue validity model (Reed, 1972; Beach, 1964) is another example of a parallel machine that predicts human classifications in at least some tasks. A weighted value is calculated independently for each dimension and then summed to determine the category assignment of the test pattern. In this case, the weights represent the uniqueness or distinctiveness of a feature as well as its frequency of occurrence within a particular category.

During the early 1960's blueprints for parallel learning machines flourished. The hope was that by adding feature detectors, more and more complex pattern spaces might be handled. Furthermore, by adding layers to the basic weighted sum device, the output of a whole set of subclassifications could be sent up as features to a higher order classification device (Nilsson, 1965). However, the early enthusiasm for perceptron models has in Hunt's (1975) words,"dampened", to say the least," since publication of the analytic treatise, Perceptrons, by Minsky and Papert in 1969. The mathematical analysis presented in that volume applied only to the simple and most basic perceptron discussed above. They reasoned that a firm mathematical understanding of the basic machine should proceed speculations regarding what more powerful versions can do. The limitations of the simple perceptron have subsequently lead Minsky and Papert (1972) to conjecture that even more complex perceptrons may lack the power necessary to handle the problems faced in human pattern recognition.

The problem of context. The crux of the argument against parallel computation machines as a model of human abilities is that they fail to deal effectively with context. They were designed to detect two dimensional figures in a context free environment quite unlike the rich, three dimensional world in which we operate. Minsky and Papert (1969) proved that while a simple perceptron can be taught to identify some object X, it will not correctly classify object X if it is placed in the context of some other object. For instance, if object X were the shape of a nose, a perceptron tuned to the shape of the nose would fail to identify it in the context of a face. Furthermore, the meaning or interpretation of some feature may depend on the context in which it is found. Since the partial predicates of a perceptron are calculated independently, context dependent meanings pose a serious problem (Hunt, 1975). An extreme example of context dependent meaning occurs in ambiguous figures, such as the Mother-in-Law-Wife illusion. The interpretation of a set of lower order features depends on the interpretation given to a number of surrounding contextual features. More commonly, the context of a scene guides the interpretation of incoming data. Reed (1975) recently reported evidence that a given set of features can be related in different manners as a function of other contextual features and relationships. It should be clear that a model, such as perceptron, that fails to deal with the hierarchical nature of real world scenes is incomplete.

Structural models are specifically designed to describe and not simply to classify the feature space. Structural approaches pick up where statistical models fail by describing a scene in terms of hierarchically related components. Each object in a scene corresponds to a well-formed expression in a "picture grammar". The grammar specifies all of the (allowable) features and relations between features which make up an object. Notice that in this framework, an object at one level of context can be a component of some higher order object. Beginning with a set of elementary primitive features and relations, one can construct a complex scene by embedding one well-formed expression with another. What is a feature and what is an object depends entirely on the level of context under consideration. At the top of the hierarchy is, of course, the entire wholistic scene--one that may be completely constructed as well as classified, by virtue of the descriptive grammar.

Palmer (Note 2) nicely described the hierarchical nature of real world concepts in the following manner (p. 5):

> As an example, consider the perceptual structure of a standing person. As a whole, the person is a rather elongated, ellipse-shaped object of some length-to-width (ratio) that is oriented vertically and has some scale or size. This might be the most global level of representation for the person, one that might be constructed from just low spatial frequency information. At a finer level of resolution the parts of the body are delineated. There is a head, torso, two arms, and two legs. Each of these parts--when considered as a whole--has global representation too. The head is a less elongated ellipsoid (of some specific length-to-width ratio) that is oriented vertically with a scaler size dependent on the size of the body. But the analysis need not stop at this level either. The head contains further parts; it has eyes, ears, a mouth, and a nose. Clearly, these parts can be represented both globally and as a further set of parts. What emerges is a multi-leveled perceptual representation of parts and wholes with obvious hierarchical structure.

A feature-relation or descriptive model of human categorization meets the problem of context-dependent meaning and the need to find an object embedded within a higher order context. Moreover, a descriptive memory system is appealing from the standpoint of theoretical economy. Many cognitive processes use or manipulate symbolic descriptions which signify our knowledge of concepts, events, objects and so on. The reader may wish to refer to Minsky and Papert (1972) for an in depth treatment of the ways in which a descriptive memory might be used in tasks other than classification. For the moment, however, a pair of examples should suffice.

Consider Winston's (1970) concept learning program. The program begins to analyze visual scenes (toy blocks world) by engaging in feature detection processes implemented by Guzman's SEE program (Minsky & Papert, 1969). The scene is divided into objects which then modify existing concept descriptions. Minsky and Papert (1969) have described the behavior of Winston's program as a sequential process which uses description matching methods and previous experience to comprehend a block world scene. Relations between features are detected and used in matching data to concept descriptions. The program learns to recognize visual concepts, such as an arch, by building a structural description which automatically classifies input as an arch if incoming data match the description. Recognizing or encoding a scene, by this approach, involves activating a set of production rules whose conditions of activation are matched by incoming data. From this, one develops a picture grammar that describes objects encountered in one's world and the description begins as a crude hypothesis and is refined as information regarding criteriality of features and relations is picked up. The rewrite rules of this grammar are presumably stored in memory to be accessed during the process of recognizing, classifying, imaging, or thinking about some visual concept or exemplar.

Another example of how descriptive methods are used in cognition comes from Evan's analogical reasoning program (see Minsky, 1968). Few would deny that the ability to apprehend and create analogies is a hallmark of human thought. The analogy problem requires that one select some scene X which best fits into the following frame: A is to B as C is to X. The relationship between A and B must be adequately described by comparing the two scenes. Next C is compared to five alternative X's (the possible answers provided in a standard visual analogies test). The different descriptions produced in the second step are compared individually with the description produced by the initial A-B comparison. The alternatives yielding the closest match in terms of common features and relations is chosen as the correct X. Again, descriptive methods lie at the heart of the program.

The emphasis on relationships among features is not new in the psychological literature (see Sutherland, 1968). For example, the basis of a feature-relation theory of concept learning can be found in a paper by Bourne (1970), who defined a concept ($\underline{C}$) in terms of its relevant features (X,Y) and some logical relationship (R):

$$\underline{C} = \underline{R}(X,Y) \qquad (1)$$

The present position is more general in that logical rules are not the only acceptable relations. Relationships in space and over time are two types that most certainly must be considered if the theory is to be applicable to real world concept formation. However, as we shall see in the next section, a theory based solely on criterial features and relations is also incomplete.

Category Organization

Most laboratory work in concept learning has used one variant or another of the rule-governed attribute identification task (Bourne, 1966). Often an affirmation rule describes the concept problem, although more complex bidimensional rules have also been studied. Assuming that one knows the rule governing the solution, the problem is one of abstracting the correct relevant feature(s). Rosch (1973) has argued that tasks of this sort are contrived in the sense that no particular member of the positive category may be considered a better example of the positive category than any other. If the conjunctive combination red-square defines the positive class, it is logically meaningless to ask which particular red-square stimulus is the best example. Irrelevant dimensions are presumed to be ignored since the rule governing the solution makes no use of them.

Natural categories, such as those denoted by nouns in language, have a different character. All members of the category are not equally good examples of the concept; instead there is a gradient of membership.

Any particular instance of a real world concept can be assigned a position along this hypothetical gradient. The peak of the gradient represents the prototype of the category. The evidence stems from six sources. The most direct approach involves a rating task. When asked to rate examples on a 7-point scale of typicality, subjects are in agreement as to how well a given instance represents the meaning of a concept. Typicality ratings have been collected for a variety of visual concepts including color, furniture, fruit, vehicle, weapon, vegetable, carpenter's too, bird, toy, and clothing categories (Rosch, 1973, 1975).

Another line of support is the finding that stimuli rating as good examples are learned faster than less prototypical items. This was found for color categories (Rosch, 1973) and artificial letter strings, dot patterns, and stick figure concepts (Rosch, Simpson, & Miller, 1976).

Third, production norms for superordinate visual concepts (Battig & Montague, 1969) tend to correlate highly with typicality ratings. Prototypical members are more likely to be given by subjects than poorer examples of the superordinate category. Mervis, Catlin, and Rosch (in press) have found that prototype norms differ from production norms in that the former are not correlated with written word frequency while the latter are. This result incidentally is consistent with an assumption underlying the present research. Namely, typicality ratings reflect our perceptual experience with real world exemplars rather than our linguistic experience with words symbolizing those exemplars. Word frequency should not be expected to bear a relation to typicality norms of visual concepts.

A fourth argument is based on a linguistic analysis of hedges. We often express degree of category membership by qualifying statements of group membership of the type an "S is a P". To illustrate, one might say, "Technically speaking, a penguin is a bird". If all members of P were equally good examples, then it would not sound peculiar to say "Technically speaking, a robin is a bird." George Lakoff initially made these observations and Rosch (1975a) incorporated his intuitions into an experimental task.

A fifth source of evidence can be found in the effects of priming on reaction time. Rosch showed that both the names of superordinate concrete categories (1975b) and basic level categories (Rosch, Mervis, Gray, Johnson & Boyes-Brian, 1976) generate expectations which facilitate encoding of good examples, but inhibit encoding of poor examples. For instance, a subject can identify as same (physical match instructions) a pair of chairs faster if the category name, furniture, precedes presentation of the pair than if no prime is provided. In contrast, a pair of poor category members, such as rugs, are actually responded to more quickly without the prime. This effect was obtained for both picture and word stimuli. Clearly, gradients of membership influence subjects performance in a task well suited to an examination of encoding processes.

The final line of evidence is related to the voluminous and complicated literature dealing with semantic categorization. The reaction time task requires subjects to verify statements of the form, "an S is a P," as quickly as possible. Smith, Shoben, and Rips (1974) argue that the notion of category organization is essential to an understanding of these data. Reaction times seem to depend on how good a member of categpry P stimulus S is (Rosch, 1973). The better the example, the faster one is at correctly verifying the statement.

A descriptive approach to concept learning cannot accommodate gradients of membership without distinguishing between features and relations of prototypical members and those of poor category members. This means that an adequate theory must be a blend of statistical and structural methods (see Reed, 1973). The mixture can be one of two types; either distance or frequency may be incorporated into the feature-relation framework. Such a combination is not uncommon. Kanal and Chandrasekaran (1972) note that the most difficult pattern recognition problems require the double-barrel approach. Although, as we have seen, descriptive methods are needed to contend with the context related problems of human classification, the phenomenon of category organization demands some distance or probability notion that allows for some features and relations to be more typical of a concept than others.

Distance versus Frequency

Either of two general statistical techniques could be injected into the pure feature-relation model. Common to both is the idea that some features are more critical to the definition of a concept than others. The first views each dimension as a continuous interval scale based on Minkowskian *r* metrics (Hunt, 1975). Statistical (Reed, 1972) or structural (Reed, 1973) models of human categorization based on city block (r =1) or Euclidean (r = 2) metrics assume that people can encode and represent a particular point on a continuous scale. In contrast, one can think of each dimension not as a continuum of points, but as a fragmented, imprecise set of intervals (Neumann, in press). Instead of storing precise values, an interval storage hypothesis claims that when a feature occurs, it is, in a sense, categorized or assigned to an interval on a dimension. The frequency distribution for each dimension is represented in memory either by means of some type of multiplexing scheme (such as a counter, a list, or an episode marker) or by a strength parameter. The point is that the complex assumptions required by metric models can be replaced by less demanding assumptions of an interval storage hypothesis.

Consider the problem in the following form. A feature-relation theory of concept acquisition and use must incorporate some statistical elements in order to account for gradients of membership. According to the prototype-distance veiw, membership is inversely related to the distance of an exemplar from the central tendancy or mean point of the multidimensional feature space. The approach assumes that absolute feature measurements made by a subject can be thought of as points on a continuum. The second approach is useful in cases where distance metrics prove unfeasible and the values of the dimensions are not points along an interval scale. The interval storage idea claims that any dimension can be chopped up into a "few" discrete values to be stored in memory. This assumption is halfway between the standard psycholinguistic assumption of binary dimensions (present-absent) and the continuum postulated by distance models. While continuous quantitative attributes allow calculation of a mean value, fragmented ordinal (or nominal) scales allow only a frequency analysis. The prototype example is made up of all the high frequency or modal dimension values.

For qualitative stimulus dimensions a non-quantitative interval analysis seems compulsory. For instance, Rosch **et al** reported prototype learning with stimuli made up of only 5-letters in which order of appearance of the elements was randomly determined (also see Reitman & Bauer, 1973; Hayes-Roth & Hayes-Roth Note). This situation involves binary dimensions--each dimension consists of but one feature which may be present or absent. As expected, the prototype consisted of the 5 letters which appeared most often in the acquisition set. With these qualitative stimulus dimensions it is difficult to imagine what a mean value on a dimension would be. Moreover, it may be unnecessary to presume that subjects ever deal with infinite values or continuous dimensions in forming concept descriptions. Our propensity to encode "chunks" of information has been well established in memory studies. If a "few" values can adequately encode the variation found even on a quantitative dimension, then there is no need to postulate a continuous representation.

Minsky (1974) has conjectured that a few qualitative intervals are sufficient to account for human visual abilities. He states the argument as follows:

> In a computer-based robot, one certainly could use metric parameters to make exact perspective calculations. But in a theory of human vision, I think we should try to find out how well our image abilities can be simulated by "qualitative" symbolic methods. People are very poor at handling magnitudes or intensities on any absolute scale; they cannot reliably classify size, loudness, pitch, weight, into even so many as ten reliably distinct categories. In comparative judgments, too, many conclusions that might seem to require numerical information are already implied by simple order, or gross order of magnitude...One thus hardly ever needs quantitative precision...

A study by Eriksen and Hake (1955) illustrates the point made by Minsky. These investigators looked at a subject's ability to distinguish between different sized squares by assigning a unique response category to each unique stimulus. The number of stimuli (5,11,21), number of responses (5,11,21), and the range of stimulus sizes (2-42 or 2-82 mm. per side) were varied between subjects. A main effect was obtained for the range variable on amount of information transmitted ($\underline{I}_t$) in bits. Discrimination was slightly better (.2 bits) for the 2-82 m.m. than for the 2-42 m.m. condition. More importantly, an interaction was found between the number of unique stimuli and the number of responses. Discrimination was impaired when there were fewer response categories than there were stimuli, but was equally good as long as there were at least as many or more responses as stimuli. That is, discrimination was equally good for subjects given 5 stimuli and 5 responses or 11 stimuli and 11 responses, or 21 stimuli and 21 responses. The values of I_t for these three groups (averaged across stimulus range) were 2.08, 2.07, and 2.08 indicating that only between four and five categories could be discriminated without error.

Absolute judgment data clearly do not support the notion that we accurately store dimension values. On the contrary, people seem to chop up size dimensions and possibly all dimensions, both quantitative as well as qualitative, into only a few values (four or five in the Eriksen and Hake study), a finding consonant with the interval storage hypothesis.

Nevertheless, the ubiquitous presence of Aristotle's Golden Mean lends considerable philosophical weight to the notion of a central tendency prototype. Moreover, there are data to suggest that subjects regard the means of quantitative dimensions to be the prototype features (Reed, 1972; Rosch, et al. 1976). How can this support for prototype-distance models be reconciled with the argument for interval-storage models?

The two types of models are difficult to disentangle experimentally, particularly with natural populations for which the dimensional probability functions are unknown. Notice that if the feature values on an attribute are normally distributed, then the mean value coincides with the model value. This may not be an uncommon situation with natural category dimensions. At least, it is difficult to say with certainty that the mode value is different from the mean value with real world concepts. To the extent that they are the same, the mean versus mode issue is moot.

Another problem poses a major obstacle to designing a decisive experiment. Without a series of extensive scaling studies it is difficult to specify the bandwidth of values or intervals on quantitative dimensions. Neumann (in press) encountered this problem in exploring one promising line of attack, here referred to as the "hole in the middle" design. He created acquisition sets in whcih the central tendency value on each of four dimensions (materials were Identikit faces and abstract geometric stimuli) was not represented; the two endpoints of the distribution occurred most often. Subjects were given a recognition test after viewing the acquisition set once, judging whether they had seen a face before and expressing their confidence in that judgment. If confidence ratings regress on frequency sums, the interval storage hypothesis model is supported. On the other hand, if the "hole in the middle" or mean feature was recognized with highest confidence, then a central tendency model is vindicated. Neumann's results were complex. Under certain conditions subjects best recognized the mean prototype, while under others, the model prototypes. For example, when subjects were not informed regarding the dimensions of variance prior to seeing the Identikit faces the former result obtained. The latter was found when subjects were so informed. Another potent factor appears to be the discriminability of the dimensional values. Using abstract geometric stimuli, Neumann gave values separated by a large interval to some subject and values which differed by half that interval to others. The high discrimination group strongly preferred the high frequency values. In contrast, the low discrimination subjects recognized all values with equal confidence. However, the "hole in the middle" predicted by frequency theory was clearly absent.

Can a reasonable explanation be given to these results? From the standpoint of a model that assumes the ability to store dimension values accurately and to calculate a mean of those values, the answer is not clear. None of the proponents of distance models have addressed these issues, although Reed (1972) does hint that with values, which are difficult to discriminate, information processing may break down, making calculation of a mean prototype unlikely. On the other hand, an interval storage hypothesis can account for the instructional effect in one of two ways. The first assumes that the features of real human faces are normally distributed. Thus, the prototype will represent the central tendency of the distribution even though it is determined by choosing the modal value. Possibly subjects not informed of the dimensions or those given poorly discriminable features imposed their real world prototype into the task environment, while informed subjects were not so biased. The latter group alone paid attention to the task specific dimensional variations by this account. A second possibility claims that subjects did not use real world prototypes, but simply formed more accurate frequency distributions when informed of dimensional variation or when given highly discriminable features. The gist of this hypothesis is that if the intervals of storage are not clearly defined, then a mean prototype can emerge even when that value never occurs during acquisition. Presumably values were not easily discriminable in Neumann's faces, yet when cued to each dimension prior to acquisition subjects were able to define the dimension values.

Neumann (in press) suggested that the width of the interval occupied by a value may overlap with the interval of an adjacent dimension value. Depending on the difficulty of discrimination, or degree of overlap among features, central values on the dimension may be incremented more often than it would be if discriminations among features were perfect. To illustrate the principle, consider the dimension of length. If the values are difficult to discriminate, the occurrence of an extremely short value may sometimes increment intervals located at the center of the dimension, but probably would never increment intervals at the other extreme end of the scale. Likewise, center values may be incremented when long values occur. Since only the center values get it from both ends, it is possible for a central tendency prototype to emerge even when the distal stimulus set never represents the mean value. Either the distance or the interval overlap hypothesis could account for the effect of informing subjects as to the dimensions of variation. Data that support a distance model, however, are less than conclusive. For instance, explanations based on the imposition of real world prototypes or the failure to discriminate dimension values during acquisition are not ruled out by either Reed's (1972) schematic face data or Rosch, et al's (1976b) schematic man (stick figures) data. In fact, a recently completed study by Chumbley, Sala & Bourne (Note 3) shifts the weight of the evidence in favor of a general frequency interpretation of human categorization. Subjects learned to structure the concept by viewing 100 sample stimuli, interspersed with blocks of typicality ratings. After each practice trial the subject was informed of the correct typicality scale value. These values were constant for both frequency and distance models for a proportion of the 81 stimuli used in the experiment (4 dimensions, 3 values). That is, predictions of the models were correlated with respect to the typicality of items given feedback during acquisition. The second stage of the task had subjects rate all 81 stimuli, without feedback. The critical data are the ratings applied to items which differentiate between models. The results strongly supported the frequency model relative to distance model. However, analyses of each subject's ratings revealed substantial individual differences. An inspection of the data coupled with post-experimental reports of the subjects lead to the conclusion that subjects are sensitive to frequency yet that sensitivity may be translated into behavior in a variety of ways. For example, dimensions were sometimes ignored or given less weight; values on dimensions that differed in frequency were sometimes treated identically. Yet, averaged over subjects and stimuli, the frequency model predominated.

Frequency or probaiblity models complement the feature-relation theory better than distance models. We may not be capable of the precision measurements required by a central tendency theory. Despite the fact that such models enjoy considerable success in artificial intelligence applications (learning machines), our knowledge of the human visual system must guide theorizing. The Chumbley, Sala & Bourne experiment taken together with Neumann's data suggest that frequency models are viable despite the fact that determination of the "bandwidth" of intervals to which frequencies apply poses a serious methodological problem. It may never be possible to specify *a priori* the dimensions and features that will be encoded during concept acquisition. These may vary in some unlawful fashion across stimulus materials, tasks, and subjects. However, the status of present research suggests that structural-frequency models are worthy of further development. The foundation of the feature-relation approach may be modified by considering the frequency or probability distribution, p(x), of values along each dimension of variance combined by the appropriate relational rules. Formally this may be expressed as follows:

$$\underline{C} - \underline{R}\,[p(x),\ p(y),\ldots] \qquad (2)$$

This formulation provides the necessary tools for dealing with human categorization of real world objects possessing some dimensional structure. Assuming that the dimensions and features can be specified, the structural frequency approach should prove useful in predicting how people will classify an item and how they organize each class internally.

Family Resemblances

Rosch et al. (1976a) also found subjects organized letter string categories, but denies that feature frequency models can adequately account for her results. She speaks instead of a family resemblance principle in which each member of the category shares one or more features in common with one or more other exemplars, but no single feature need be common to all examples. The dimensions manipulated in her experiment, which is the limiting case of a dimension, were merely binary; A feature was either present or it was absent. Each stimulus consisted of five dimensions per stimulus (five letter strings) and each of two categories contained six stimuli. In Experiment 1 subjects learned to classify the 12 stimuli to a criterion of two errorless runs. When a subject reached criterion, he continued to classify the items in a reaction time situation for 15 additional runs. No new items were tested--only the 12 learning exemplars. Next a prototype rating and exemplar reconstruction task were administered. As expected, the rated typicality of a letter string depended on the extent to which constituent letters were common among the other five stimuli in the category. Rosch calls the operative principle family resemblance, but it appears to be isomorphic to a feature frequency model. First, consider her description of how a family resemblance score was assigned to each stimulus: "...each letter received a weight (1-5) representing the number of strings in the category in which it occurred; the weights of each letter in a string were then summed to generate the family resemblance score of that string." Replace the word "weight" with frequency score and reread. Neumann (Note 1) has recently noted the connection between Rosch's family resemblance model and feature frequency.

We doubt that Rosch would agree with the above analysis since she pointedly rejects a feature frequency interpretation of these data. Here is the rationale.

> "...an attribute frequency model (Reitman & Bower, 1973; Neumann, 1974) will not account for all of our results even for the family resemblance stimuli. In Experiment 2, typicality remainded a function of category structure even when frequencies inverse to structural typicality eliminated attribute frequency from the learning set."

The manipulation referred to in the last sentence should be evaluated carefully. In one experiment, Rosch et al. found that strings with high frequency scores were learned in fewer errors than poor category members. To equate degree of learning on each item in a second experiment, the number of times an item was presented for study was made proportional to the mean number of trials taken to learn it in the first study. This manipulation drastically alters the <u>absolute</u> frequency of the various features; however, it is not clear that the relative frequencies of values would change as Rosch et al. claim. That is, the most frequent letters may have been the same in both experiments. Although it is not possible to ascertain from their report precisely how often each string was presented, an analysis of the stimulus sets make it difficult to see how the <u>order</u> of feature frequencies could have changed. Each additional occurrence of a poor exemplar increases the frequency of idiosyncratic features. However, since all items possess at least one prototype feature, the features of the prototype are incremented when poor members are presented. When one adds in the frequencies contributed by presentations of the remaining category members, which share still more features in common, it would seem that the prototype is no different than it is under conditions of equal presentation of all exemplars. In fact, given that Rosch's explanation appears to be a special case of a structural-frequency model, it would be paradoxical if the above line of reasoning did not apply to the Rosch et al. manipulation.

Despite the apparent similarity between the empirical models of family resemblance and summed feature frequency, one finds a divergence at a deeper level. The thrust of family resemblance is the denial of defining properties of a concept that are common to all instances. Whereas feature frequency models imply the existence of at least one high frequency characteristic shared by most members of a class, family resemblance makes the claim that the members are related in a much more indirect manner. For example, the items AB, BC, and CD share a family resemblance despite the fact that no single element is common to all three.

Rosch et al. (1976 b) contend that superordinate categories are excellent examples of the family resemblance principle. As before, however, the data are open to a structural-frequency interpretation. Experiment 1 in the Rosch et al. (1976 b) paper involved the collection of attribute listings for the hypothesized subordinate (kitchen chair), basic (chair), and superordinate (furniture) levels. Each subject listed attributes of nine items; each item belonged to a unique category, yet all items were either subordinate, basic, or superordinate concepts. For instance, subjects assigned to the superordinate condition listed attributes which they felt applied to members of each of the following nine categories: musical instrument, fruit, tool, clothing, furniture, vehicle, trees, fish, and bird. The listed attributes were classified by the experimenter as noun attributes, such as "legs"; adjective attributes, such as "four-legs"; and general characteristics, such as "you sit on it".

The results for the non-biological categories will be described here since the assumed superordinate levels of fish, trees, and bird showed a large number of attributes in common, unlike the other six superordinates. Superordinates showed few dimensions common to all members and 70% of those listed were functional attributes of a general nature (you eat it, you sit on it). The basic level showed significantly more attributes in common. The subordinate level apparently is even more specific than the basic level; not only does it have more dimensions in common, but also the particular features present on each dimension tend to be more homogeneous. Note that although a small difference existed between the common attribute means of basic and subordinate levels, nearly all of the additional listings were adjectives. For instance, in addition to the noun attributes listed for the basic level chair (legs, back, seat), a living room chair drew more specific feature listings: it was large, soft, and cushioned.

A central relation among features is the logical rule governing their combination (Bourne, 1970). The common structurally-related dimensions of basic and subordinate concepts are conjunctively related. Thus, it is not surprising that many common attributes were listed for these levels. The major difference between the two levels appears to be that the latter type share more specific features in common than the former. The conjunctive relation can be illustrated with our concept of a face; an object is not a face unless a nose AND a mouth AND a pair of eyes are present. Of course, particular spatial relations must also be met, but the logical relation alone allows one to differentiate between low levels and the superordinate level of categrization. The elements of a superordinate can be thought of as a variety of basic level objects that are related disjunctively to one another. The appearance of either one OR another OR several of the basic features is sufficient to allow classification at the top level. The disjunctive rule explains why superordinates share few physical dimensions. While one may label this situation as an instance of family resemblance principles, it clearly falls in the domain of the structural-frequency theory. The physical attributes are conjunctively related at the basic level, then each basic object (e.g., cars, trucks, bicycles) serves as a higher order feature of the disjunctively related superordinate (vehicles). The common feature at the superordinate level is some general or functional characteristic as Rosch's results indicate. (A vehicle is a means of transporting something.) Somewhat at variance with this assertion are results reported by Rosch and Mervis (1975) showing that the least typical members of the fruit, weapon, vegetable, and clothing categories were not viewed by subjects as sharing any attributes in common. It is not clear why subjects failed to mention the defining, albeit general, characteristic of these categories, although olives, like other fruits, can be eaten, a screwdriver can be used to fight with, rice can be eaten, and a necklace can be worn.

The results are possibly due to a procedural problem. Subjects were not informed what category they should consider the test item a member of in listing its properties. This would pose no problem for good instances of the concept, but it is doubtful that the defining property of all weapons would come to mind when one is asked to describe a screwdriver. At any rate, it is inappropriate to conclude that defining features do not exist on the basis of a test which may not adequately tap the subject's knowledge. This is especially true in the case of functional features. Although a person may never state that a screwdriver can be used to fight with, he may make use of that property in an appropriate context.

Feature Extraction

In terms of physical, observable characteristics superordinate concepts lack a sinqle defining property, although common characteristics can be named in a functional domain. Unfortunately, it is never clear what the appropriate feature description should be. We have assumed in our work using facial stimuli that the "obvious" features we manipulated were the ones encoded by the subject. Perhaps the subjects parsed the face quite differently, keeping track of such abstract properties as the "happiness of the facial expression." Certainly some of the individual differences and failures to fit the summed frequency model can be traced to our inability to specify which features the subject counted. Indeed the main weakness of the structural-frequency position, or any feature theory for that matter, is ambiguity surrounding the issue of feature extraction (Hunt, 1975).

The appropriate description or representation of the feature-relation space for real world concepts is an important area of research. Palmer's (1975) implementation of the Gestlt principles of grouping is one example of the type of research that needs to be performed. It may be the case that the obvious, intuitive features used to communicate the structural-frequency point of view have nothing to do with the features actually involved. Furthermore, the best representation of the features may be something quite unlike anything we might intuit. For instance, the visual system may perform a Fourier transformation on the proximal stimulus and then carry out feature matching procedures in the Fourier domain (Duda & Hart, 1972). The proper level of description may be far more abstract than the "obvious" features we consciously perceive. The claim is that at some level of feature description the structural-frequency theory outlined here is of use in describing human categorization.

Summary

In the present paper we reviewed theories of pattern recognition which can account for a person's ability to classify and organize real world concepts according to a dimension of typicality. Only theories based on a feature format of mental representations were considered. One difference between models is whether relations between features are critical or not. Although models of semantic memory emphasize the importance of relations between objects (e.g., Anderson & Bower, 1973; Kintsch, 1975, Rosch & Mervis, 1975), pattern recognition models, including family resemblance (Rosch & Mervis, 1975) commonly operate on a list of independent features. It was argued that a relational or structural approach is necessary to handle the contextual problems encountered in human pattern recognition.

Nevertheless a purely structural approach is also insufficient. A vari of lines of evidence indicate that natural categories are internally organized according to typicality. Thus, either a structural-distance model or a structural-frequency model is called for, with the statistical component accounting for gradients of membership (see Reed, 1973). Although the empirical support for one approach over the other is scanty, it was argued that the frequency approach is preferable.

People do not seem to be capable of the precise level of feature measurement displayed by successful computer models of pattern recognition. It is more plausible that one encodes dimensions in a rough, qualitative fashion rather than as points on a continuous scale of measurement. It is already well known that summed frequency is useful in predicting typicality ratings in the case of qualitative dimensions not amenable to the distance approach. Rosch et al. (1976) suggest that a person uses both frequency and distance pattern recognition methods depending on whether the dimensions of variance are better characterized in quantitative or qualitative terms. Because there is no strong evidence to the contrary, however, we attempt to account for both situations with the same theory.

It would appear that some form of structural-frequency theory is powerful enough to cope with the problems in human pattern recognition. The family resemblance principle offered by Rosch and Mervis (1975) was shown to be a special case of such a theory. Family resemblance describes the nature of the physical properties shared by instances of superordinate categories. For superordinates, relations between features may be unimportant, but as a general theory, family resemblance would fail without the inclusion of relation information. The claim that superordinates share no defining feature was questioned. At the level of general, functional characteristics superordinates can be described in terms of dimensional frequency information just as well as basic and subordinate categories. The family resemblance issue should serve to sensitize us to the pervasive problem of specifying the proper feature space. It remains to be seen whether the appropriate features for human categorization can be decided upon.

Reference Notes

1. Neumann, P. G. Family resemblance to the attribute frequency model: A chip off the old block? Manuscript submitted for publication, 1977.

2. Palmer, S. E. Hierarchical structure in perceptual representation. Manuscript submitted for publication.

3. Chumbley, J. I., Sala, L. S., & Bourne, L. E., Jr. The bases of acceptability ratings in quasi-naturalistic concept tasks. Manuscript submitted for publication, 1976.

References

Anderson, J. R., & Bower, G. H. Human Associative Memory. Washington, D.C. Winston, 1973.

Andrews, H. Introduction to mathematical techniques in pattern recognition. Toronto: Wiley, 1972.

Battig, W. G., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. Journal of Experimental Psychology Monograph, 1969, 80, (3, Pt 2).

Bourne, L. E., Jr. Knowing and using concepts. Psychological Review, 1970, 77, 546-556.

Bourne, L. E., Jr. Human conceptual behavior. Boston: Allyn and Bacon, 1966.

Duda, R. & Hart, P. Pattern classification and scene analysis. Toronto: Wiley, 1973.

Erikson, C & Hake, H. Absolute judgments as a function of the stimulus range and the number of stimulus and response categories. Journal of Experimental Psychology, 1955, 49, 323-332.

Franks, J. & Bransford, J. Abstraction of visual patterns. Journal of Experimental Psychology, 1971, 90, 65-74.

Hunt, E. B. Artificial intelligence. New York: Academic Press, 1975.

Kanal, L. & Chandrasedaran, B. On linguistic, statistical, mixed models for pattern recognition. In Watanabe, S. (Ed.) Frontiers of pattern recognition London: Academic Press, 1972.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Mervis, C.B., Catlin, J., & Rosch, E. Relationship among goodness-of-example, category norms, and word frequency. Bulletin of the Psychonomic Society, in press.

Minsky, M. (Ed) Semantic information processing, Cambridge: MIT Press, 1968.

Minsky, M. A framework for representing knowledge. Artificial Intelligence Memo No. 306, MIT AI Laboratory, Cambridge, Mass., June, 1974.

Minsky, M. & Papert, S. Perceptrons, Cambridge: MIT Press, 1969.

Minsky, M. & Papert, S. Progress Report on Artificial Intelligence. Artificial Intelligence Memo. No. 306, MIT AI Laboratory. Cambridge, Mass., Jan., 1972.

Neumann, P. G. Visual prototype formation with discontinuous representation of dimensions of variability. Memory & Cognition, in press.

Nilsson, N. Learning machines. New York: McGraw-Hill, 1965.

Reed, S. K. Pattern recognition and categorization. Cognitive Psychology, 1972, 3, 382-407.

Reed, S. K. Psychological Processes in Pattern Recognition. New York: Academic Press, 1973.

Rosch, E. On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press, 1973.

Rosch, E. The nature of mental codes for color categories. Journal of Experimental Psychology: Human Perception and Performance, 1975a, 1, 303-322.

Rosch, E. Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 1975b, 104, 192-233.

Rosch, E. & Mervis, C. B. Family resemblances: Studies in the internal structure of categories, Cognitive Psychology, 1975, 7, 573-605.

Rosch, E., Simpson, S., & Miller R. Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 1976a, 4, 491-502.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Brian, P. Basic objects in natural categories. Cognitive Psychology, 1976b, 8, 382-439.

Sutherland, N. Object recognition. In Carterette, E. & Friedman, M. Handbook of perception, Vol. III, Biology of Perceptual System. New York: Academic Press, 1973.

Winston, P. H. Learning structural descriptions from examples. Artificial Intelligence Memo No. TR-231, MIT AI Laboratory, Cambridge, Mass., September, 1970.